

In the Specification:

On page 22, please amend the paragraph beginning at line 21 as follows:

--Genomic signal processing is often concerned with finding DNA (deoxyribonucleic acid) sequences on longer strands of DNA consisting of the adenosine, thymine, cytosine or guanine (A, T, C and G) nucleotides. Thus a DNA strand can be described as a sequence of the symbols A, T, C and G. Identical, but also similar DNA sequences which differ at certain positions can be found by the above mentioned robust search algorithm. This similarity comparison is even more pronounced in the case that the sequences are long, as is typically the case in DNA analysis with thousands of nucleotides. The application of the present invention to the genome sequence identification is a good example for a search algorithm wherein the fragmentation has nearly no effect, as the low number of entities (4) makes it improbable, that a sequence occurs ~~metnot~~ having a nearly even distribution of nucleotides. The Genome analysis on the other hand discloses aspect of the entities, instead of the 4 nucleotides itself, the 20 nucleotide-coded amino acids can be selected to describe and compare the sequences, wherein one entity is represented by one or more 3-tuples of nucleotides. It may be noted that in the case of genome analysis on the basis of the DNA, the speed up techniques like fragmentation and entity recognition can not be applied, as it is very likely that e.g. in a 200 entity sequence, the single nucleotides are more or less evenly distributed. So it can be assumed that fragmentation techniques can not help to speed up the recognition process. In the case of nucleotide based string analysis, it would also be not applicable to provide a toleration of a faulty value for the single nucleotides, because it is not clear, which one is the "nearest" nucleotide for e.g. Tymine. In case of an RNA triplet analysis, the nucleotide triplets are each coding an amino acid. The triplets can provide a sequence as $4*4*4=64$ triplets are coding only 20 different amino acids or commands. In this case, it is clear that e.g. UUA, UUG, CUU, CUC, CUA and CUG code the same amino acid leucin and can therefore be regarded as equal. To apply the present invention on these triplets, the pairs of amino acids or triplets are chosen like the following (GAU, UAC) defining the order of Aspargin followed by Tyrosine or as the pair (asp, tyr) to be compared in two RNA strings.

Speech recognition (see Figure 10).--